REVIEW

In $\mathbb{R}^n$: $\mathcal{B} = \{\boldsymbol{u_1}, ..., \boldsymbol{u_i}, ...\boldsymbol{u_p}\}$

$\mathcal{B}$ is an orthogonal set means $\boldsymbol{u_i} \perp \boldsymbol{u_j}$ for all $i \neq j$

$\mathcal{B}$ is an orthonormal set means $\mathcal{B}$ is orthogonal and $||\boldsymbol{u_i}|| = 1$ for each i

If $\mathcal{B} = \{\boldsymbol{u_1}, ..., \boldsymbol{u_i}, ..., \boldsymbol{u_p}\}$ = an **orthogonal basis** for a subspace W of $\mathbb{R}^n$

For $\boldsymbol{y} \in W$:

$$\boldsymbol{y} = \underset{\downarrow}{c_1}\ \boldsymbol{u_1} + ... + \underset{\downarrow}{c_i}\ \boldsymbol{u_i} + ... + \underset{\downarrow}{c_p}\ \boldsymbol{u_p}$$

$$= \frac{\boldsymbol{y}\cdot\boldsymbol{u_1}}{\boldsymbol{u_1}\cdot\boldsymbol{u_1}}\boldsymbol{u_1} + ... + \frac{\boldsymbol{y}\cdot\boldsymbol{u_i}}{\boldsymbol{u_i}\cdot\boldsymbol{u_i}}\boldsymbol{u_i} + ... + \frac{\boldsymbol{y}\cdot\boldsymbol{u_p}}{\boldsymbol{u_p}\cdot\boldsymbol{u_p}}\boldsymbol{u_p} \qquad (*)$$

If $\mathcal{B} = \{\boldsymbol{u_1}, ..., \boldsymbol{u_i}, ..., \boldsymbol{u_p}\}$ = **orthonormal** basis for a subspace W of $\mathbb{R}^n$ (so each $\boldsymbol{u_i} \cdot \boldsymbol{u_i} = 1$)

$$\boldsymbol{y} = (\boldsymbol{y}\cdot\boldsymbol{u_1})\,\boldsymbol{u_1} + ... + (\boldsymbol{y}\cdot\boldsymbol{u_i})\,\boldsymbol{u_i} + ... + (\boldsymbol{y}\cdot\boldsymbol{u_p})\,\boldsymbol{u_p}$$

A nicer formula, but using a normalized basis in arithmetic calculations may be messier: see the following example

Example

$\mathcal{B} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}$, an orthogonal basis for $W = \mathbb{R}^3$.

To write $\boldsymbol{y} = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$

Find $c_1 = \dfrac{\begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}$, $c_2 = \dfrac{\begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}}{\begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}}$, $c_3 = \dfrac{\begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}}{\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}}$

$c_1 = \frac{8}{2} = 4$ $\quad c_2 = -\frac{2}{18} = -\frac{1}{9}$ $\quad c_3 = -\frac{14}{9}$

so $\begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{9} \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix} - \frac{14}{9} \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$

If $\mathcal{B} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \end{bmatrix}, \begin{bmatrix} \frac{2}{\sqrt{9}} \\ \frac{1}{\sqrt{9}} \\ \frac{-2}{\sqrt{9}} \end{bmatrix} \right\}$, an orthonormal basis for $\mathbb{R}^3$

from normalizing the preceding basis $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}$

Then $\boldsymbol{y} = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + c_3\boldsymbol{u_3}$, where

$c_1 = \boldsymbol{y} \cdot \boldsymbol{u_1} = \begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{8}{\sqrt{2}}$

$c_2 = \boldsymbol{y} \cdot \boldsymbol{u_2} = \frac{-2}{\sqrt{18}}$ and $c_3 = \boldsymbol{y} \cdot \boldsymbol{u_3} = \frac{-14}{\sqrt{9}}$ and

$\begin{bmatrix} 1 \\ -2 \\ 7 \end{bmatrix} = \frac{8}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \frac{2}{\sqrt{18}} \begin{bmatrix} \frac{-1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \end{bmatrix} - \frac{14}{\sqrt{9}} \begin{bmatrix} \frac{2}{\sqrt{9}} \\ \frac{1}{\sqrt{9}} \\ \frac{-2}{\sqrt{9}} \end{bmatrix}$

Example (*the background theory and formulas discussed in the lecture and textbook*)

Let $L =$ Span $\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$. Write $\boldsymbol{y} = \begin{bmatrix} 5 \\ 0 \\ -3 \end{bmatrix}$ as the sum of a vector from L and a vector from $L^{\perp}$. Find the perpendicular distance from $\boldsymbol{y}$ to L.

$$\widehat{\boldsymbol{y}} = \text{proj}_L \boldsymbol{y} = \frac{\begin{bmatrix} 5 \\ 0 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$= \frac{2}{6} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}. \quad \text{Then}$$

$$\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{bmatrix} 5 \\ 0 \\ -3 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{14}{3} \\ \frac{2}{3} \\ -\frac{10}{3} \end{bmatrix} \quad (\textit{note: } \perp \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \textit{ a check})$$

$$\begin{bmatrix} 5 \\ 0 \\ -3 \end{bmatrix} = \boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z} = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{14}{3} \\ \frac{2}{3} \\ -\frac{10}{3} \end{bmatrix}$$

1) $\widehat{\boldsymbol{y}}$ is the point on L closest to $\boldsymbol{y}$

2) The distance from $\boldsymbol{y}$ to L is $||\boldsymbol{z}|| = ||\boldsymbol{y} - \widehat{\boldsymbol{y}}||$

$$= \sqrt{\tfrac{196}{9} + \tfrac{4}{9} + \tfrac{100}{9}} = \sqrt{\tfrac{300}{9}} = \tfrac{10\sqrt{3}}{3}$$

3) $||\boldsymbol{y} - \widehat{\boldsymbol{y}}|| = \frac{10\sqrt{3}}{3}$ is the error if $\widehat{\boldsymbol{y}}$ is used as an approximation for $\boldsymbol{y}$

4) $\widehat{\boldsymbol{y}}$ is the "best approximation to $\boldsymbol{y}$ from L" (meaning: if some other $\boldsymbol{w}$ from L is used to approximate $\boldsymbol{y}$, then the error $||\boldsymbol{y} - \boldsymbol{w}|| > \frac{10\sqrt{3}}{3}$

Example $\mathcal{B} = \{\boldsymbol{v_1}, ..., \boldsymbol{v_i}, ...\boldsymbol{v_n}\}$ = an orthogonal basis for subspace W of $\mathbb{R}^n$

For $\boldsymbol{y} \in W$: This is

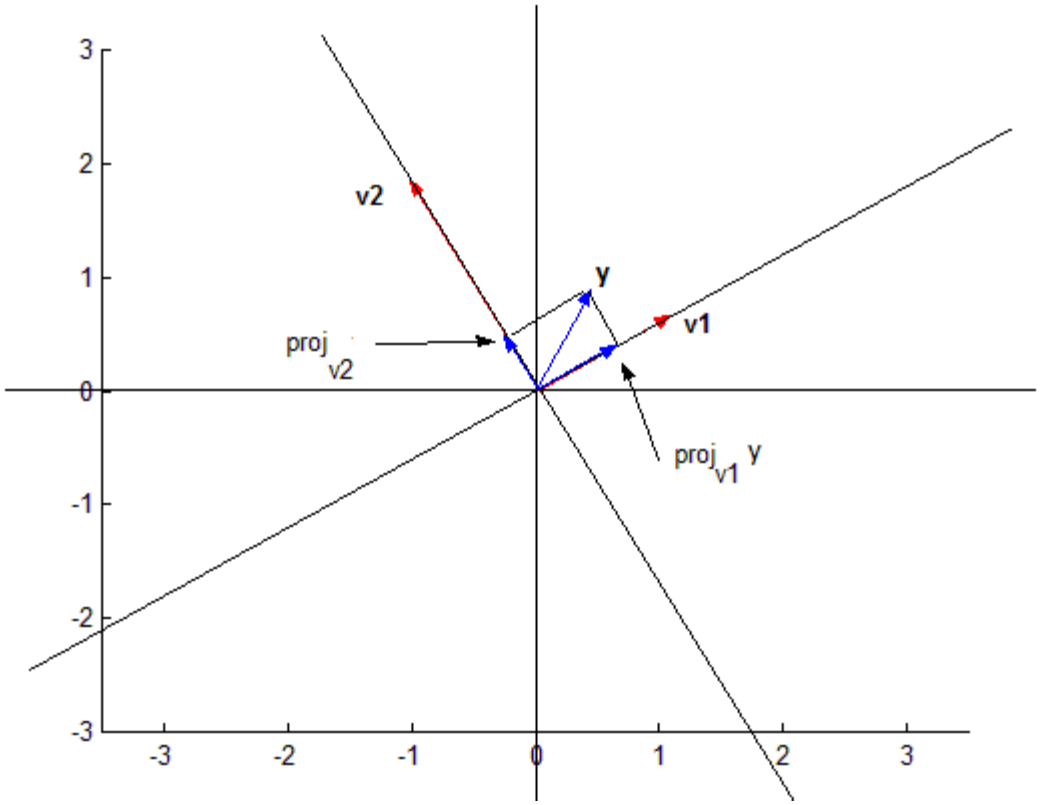
$$\boldsymbol{y} = \frac{\boldsymbol{x}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + ... + \frac{\boldsymbol{x}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}}\boldsymbol{v_i} + ... + \frac{\boldsymbol{x}\cdot\boldsymbol{v_n}}{\boldsymbol{v_n}\cdot\boldsymbol{v_n}}\boldsymbol{v_n} \quad \leftarrow \text{This is (*) from the beginning of these notes}$$

$$= \text{proj}_{\boldsymbol{v_1}}\boldsymbol{y} + ... + \text{proj}_{\boldsymbol{v_i}}\boldsymbol{y} + ... + \text{proj}_{\boldsymbol{v_n}}\boldsymbol{y} \quad \leftarrow \text{rewritten using projections}$$

See illustration below for $\mathbb{R}^2$

In $\mathbb{R}^2$: for an orthogonal basis $\mathcal{B} = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$

$$\boldsymbol{y} = \frac{\boldsymbol{x}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + \frac{\boldsymbol{x}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2} = \text{proj}_{\boldsymbol{v_1}}\boldsymbol{x} + \text{proj}_{\boldsymbol{v_2}}\boldsymbol{x}$$



Theorem An $m \times n$ matrix A has orthonormal columns if and only if $U^T U = I$

Note: there is no particular special name for this kind of matrix.)

Proof *(Just look at 3×3 case – only to make the notation look simpler)*

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}$$

$$U = [\, \boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \boldsymbol{u_3} \,], \text{ so that } U^T = \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix}.$$

$$U^T U = \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix} [\, \boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \boldsymbol{u_3}]]$$

$$= [\, \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix} \boldsymbol{u_1} \quad \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix} \boldsymbol{u_2} \quad \begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix} \boldsymbol{u_3} \,]$$

But

$$\begin{bmatrix} \boldsymbol{u_1}^T \\ \boldsymbol{u_2}^T \\ \boldsymbol{u_3}^T \end{bmatrix} \boldsymbol{u_1} = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u_1^T u_1} \\ \boldsymbol{u_2^T u_1} \\ \boldsymbol{u_3^T u_1} \end{bmatrix}$$

and in general

$$U^T U = \begin{bmatrix} \boldsymbol{u_1^T u_1} & \boldsymbol{u_1^T u_2} & \boldsymbol{u_1^T u_3} \\ \boldsymbol{u_2^T u_1} & \boldsymbol{u_2^T u_2} & \boldsymbol{u_2^T u_3} \\ \boldsymbol{u_3^T u_1} & \boldsymbol{u_3^T u_2} & \boldsymbol{u_3^T u_3} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{u_1 \cdot u_1} & \boldsymbol{u_1 \cdot u_2} & \boldsymbol{u_1 \cdot u_3} \\ \boldsymbol{u_2 \cdot u_1} & \boldsymbol{u_2 \cdot u_2} & \boldsymbol{u_2 \cdot u_3} \\ \boldsymbol{u_3 \cdot u_1} & \boldsymbol{u_3 \cdot u_2} & \boldsymbol{u_3 \cdot u_3} \end{bmatrix}$$

so $U^T U = I$ if and only if

$$\boldsymbol{u_1 \cdot u_1 = u_2 \cdot u_2 = u_3 \cdot u_3 = 1}$$

and $\boldsymbol{u_i \cdot u_j} = 0$ if $i \neq j$

Theorem Suppose U is $m \times n$ and that U has orthonormal columns. Multiplication$U$ defines a linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m : \quad T(\boldsymbol{x}) = U\boldsymbol{x}$

Then for $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$

1) $\boldsymbol{x} \cdot \boldsymbol{y} = (U\boldsymbol{x}) \cdot (U\boldsymbol{y})$ (the transformation preserves inner products)

2) $||\boldsymbol{x}|| = ||U\boldsymbol{x}||$ (the transformation preserves length of vectors products)

Proof 1) Use the definition of inner product as a matrix product, and properties of transposes:

inner product *matrix product*

$$(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = (U\boldsymbol{x})^T(U\boldsymbol{y}) = \boldsymbol{x}^T U^T U\boldsymbol{y} = \boldsymbol{x}^{\boldsymbol{T}} I\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$$

2) $||U\boldsymbol{x}||^2 = (U\boldsymbol{x}) \cdot (U\boldsymbol{x}) = (U\boldsymbol{x})^T(U\boldsymbol{x}) = \boldsymbol{x}^T U^T U\boldsymbol{x}$

$$= \boldsymbol{x}^{\boldsymbol{T}} I\boldsymbol{x} = \boldsymbol{x} \cdot \boldsymbol{x} = ||\boldsymbol{x}||^2, \text{ so } ||\,U\boldsymbol{x}\,|| = ||\,\boldsymbol{x}\,|| \quad \bullet$$

Notice that we can then draw a couple of other conclusions:

3) $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ if and only if (by part 1) of the theorem) $U\boldsymbol{x} \cdot U\boldsymbol{y} = 0$

so $\boldsymbol{x} \perp \boldsymbol{y}$ if and only if $U\boldsymbol{x} \perp U\boldsymbol{y}$

This says that the transformation $T(x) = Ux$ preserves orthogonality

4) By a formula in the previous lecture:

$\boldsymbol{x} \cdot \boldsymbol{y} = ||\,\boldsymbol{x}\,||\,||\,y\,|| \cos\theta$ where θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$

and $U\boldsymbol{x} \cdot U\boldsymbol{y} = ||\,U\boldsymbol{x}\,||\,||\,Uy\,|| \cos\alpha$ where α is the angle between $U\boldsymbol{x}$ and $U\boldsymbol{y}$

Since $\boldsymbol{x} \cdot \boldsymbol{y} = U\boldsymbol{x} \cdot U\boldsymbol{y}$ and $||\,\boldsymbol{x}\,||\,||\,U\boldsymbol{x}\,||$ and $||\,y\,|| = ||U\,y\,||$

we conclude that $\cos\theta = \cos\alpha$ and (since both α and θ are between 0 and π) therefore $\theta = \alpha$ ()

This says that the transformation $T(\boldsymbol{x}) = U\boldsymbol{x}$ preserves the angle between any two vectors (and therefore, as already mentioned in 3). it preserves right angles. So statement 3) is "included within" the statement in .4